

A KRIGING PROCEDURE FOR PROCESSES INDEXED BY GRAPHS

T. ESPINASSE & J-M. LOUBES

INSTITUT DE MATHÉMATIQUES DE TOULOUSE, FRANCE

ABSTRACT. We provide a new kriging procedure of processes on graphs. Based on the construction of Gaussian random processes indexed by graphs, we extend to this framework the usual linear prediction method for spatial random fields, known as kriging. We provide the expression of the estimator of such a random field at unobserved locations as well as a control for the prediction error.

Keywords: Gaussian process, graphs, kriging.

INTRODUCTION

Data presenting an inner geometry, among them data indexed on graphs are growing tremendously in size and prevalence these days. High dimensional data often present structural links that can be modeled through a graph representation, for instance the World Wide Web graph or the social networks well studied in history or geography [9], or molecular graphs in biology or medicine for instance. The definition and the analysis of processes indexed by such graphs is of growing interest in the statistical community. In particular, the definition of graphical models by J.N. Darroch, S.L. Lauritzen and T.P. Speed in 1980 [5] fostered new interest in Markov fields, and many tools have been developed in this direction (see, for instance [11] and [10]). When confronted to missing data or to forecast the values of the process at unobserved sites, many prediction methods have been proposed in the statistics community over the past few years. Among them, we focus in this paper on a new method extending the kriging procedure to the case of a Gaussian process indexed by a graph.

Kriging is named for the mining engineer Krige, whose paper [8] introduced the method. For background on kriging see [1] or [3]. Gaussian process models also called Kriging models are often used as mathematical approximations of expensive experiments. Originally presented in spatial statistics as an optimal linear unbiased predictor of random processes, its interpretation is usually restricted to the convenient framework of Gaussian Processes (GP). It is based on the computation of the

conditional expectancy, which requires a proper definition of the covariance structure of the process. For this, we will use the spectral definition of a covariance operator based on the adjacency operator. Inspired by [7], we define Gaussian fields on graphs using their spectral representation (see for instance in [2]). Then, we extend to this case the usual method of prediction using the Kriging method. For this we will use blind prediction technics generalizing to graphs process the procedure for time series presented in [6].

The paper falls into the following parts. Section 1 is devoted to the definitions of a graph and the presentation of the problem. The forecast problem is tackled in Section 2. The proofs are postponed to the Appendix.

1. PRELIMINARY NOTATIONS AND DEFINITIONS

1.1. Random processes indexed by graphs. In this section, we introduce both the context and the objects considered in the whole paper.

First, assume that \mathbf{G} is a graph, that is a set of vertices G and a set of edges $E \subset G \times G$. In this work, G is assumed to be infinite (but countable).

Two vertices $i, j \in G$ are neighbors if $(i, j) \in E$. The degree $d(i)$ of a vertex $i \in G$ is the number of neighbors and the degree of the graph \mathbf{G} is defined as the maximum degree of the vertices of the graph \mathbf{G} :

$$\deg(\mathbf{G}) := \max_{i \in G} \deg(i).$$

From now on, we assume that the degree of the graph \mathbf{G} is bounded, that is

$$\exists d_{\max} > 0, \forall i \in G, d(i) \leq d_{\max}.$$

Furthermore, the graph \mathbf{G} is endowed with the natural distance d_G , that is the length of the shortest path between two vertices.

In the following, we will consider the renormalized adjacency operator A of \mathbf{G} . So its entries belong to $[-\frac{1}{\deg(\mathbf{G})}, \frac{1}{\deg(\mathbf{G})}]$. It is defined as

$$A_{ij} = \frac{1}{d_{\max}} \mathbb{1}_{i=j}.$$

We denote by B_G the set of all bounded Hilbertian operators on $l^2(G)$ (the set of square sommable real sequences indexed by G).

To introduce the spectral decomposition, consider the action of the adjacency operator on $l^2(G)$ as

$$\forall u \in l^2(G), (Au)_i := \sum_{j \in G} A_{ij} u_j, (i \in G).$$

The operator space B_G will be endowed with the classical operator norm

$$\forall T \in B_G, \|T\|_{2,op} := \sup_{u \in l^2(G), \|u\|_2 \leq 1} \|Tu\|_2,$$

where $\|\cdot\|_2$ stands for the usual norm on $l^2(G)$.

Notice that, as the degree of \mathbf{G} and the entries of A are both bounded, A lies in B_G , and we have

$$\|A\|_{2,op} \leq 1.$$

Finally we get that A is a symmetric bounded normal Hilbertian operator. Recall that for any bounded Hilbertian operator $A \in B_G$, the spectrum $\text{Sp}(A)$ is defined as the set of all complex numbers λ such that $\lambda \text{Id} - A$ is not invertible (here Id stands for the identity on $l^2(G)$). So the spectrum of the normalized adjacency operator A is a non-empty compact subset of \mathbb{R} .

Using the spectral representation of the graph, we proved in [7] that we can define Gaussian processes indexed on graphs whose covariance structure relies only on the geometry of the graph via its spectrum. For this, for any bounded positive function f , analytic on the convex hull of $\text{Sp}(A)$, note first that $f(A)$ defines a bounded positive definite symmetric operator on $l^2(G)$. Then we can define a Gaussian process $(X_i)_{i \in G}$ indexed by the vertices G of the graph \mathbf{G} with covariance operator Γ defined as

$$\Gamma = \int_{\text{Sp}(A)} f(\lambda) dE(\lambda).$$

Such graph analytical processes extend the notion of time series to a graph indexed process. So using this terminology, we will say that X is

- MA_q if f is a polynomial of degree q .
- AR_p if $\frac{1}{f}$ is a polynomial of degree p which has no root in the convex hull of $\text{Sp}(A)$.
- $ARMA_{p,q}$ if $f = \frac{P}{Q}$ with P a polynomial of degree p and Q a polynomial of degree q with no roots in the convex hull of $\text{Sp}(A)$.

Otherwise, we will talk about the MA_∞ representation of the process \mathbf{X} . We call f the *spectral density* of the process \mathbf{X} , and denote its corresponding covariance operator by

$$\Gamma = \mathcal{K}(f) = f(A)$$

So the spectral analysis of the graph enables to define a class of admissible covariances for stationary Gaussian processes with associated spectral density f . Hereafter we tackle the issue of prediction of such processes.

1.2. Blind prediction problem. We can now introduce our prediction problem on this framework.

The problem comes from practical issues. In real life problems, it is usual to own a single sample, which has to be used for both estimation and prediction.

Let f be a bounded positive function, analytic over $\text{Sp}(A)$, and $(X_i)_{i \in G}$ be a Gaussian zero-mean process indexed by G of covariance operator $\mathcal{K}(f)$.

We will observe the process on a growing sequence of subgraphs of G , but with missing values we aim at predicting. Let (O, B) be a partition of G . The set O will denote the set of indexes for all the possibly observed values while B denotes the "blind" missing values index set. In all the following, the set B where the observations should be forecast, is assumed to be finite.

Let $(G_N)_{N \in \mathbb{N}}$ be a growing sequence of induced subgraphs of G . This means that we have at hand a growing sequence of vertices and consider as the observed graph all the existing edges between the vertices that are observed. From now on, we assume that N is large enough to ensure $B \subset G_N$. The observation index set will be denoted $O_N := O \cap G_N$.

Hence, we consider the restriction $X_{O_N} := (X_i)_{i \in O_N}$, which stands for the data we have at hand at step N . We consider the asymptotic framework where the observations O_N fill the space between the blind part and the graph, in the sense that the distance between the blind locations B and the non observed graph $G \setminus O_N$ increases, i.e

$$m_N := \frac{1}{d_G(B, G \setminus O_N)} \longrightarrow 0, \quad \text{when } N \rightarrow +\infty.$$

This corresponds to the natural case where the blind part of the graph becomes more and more surrounded by the observations without any gap. So the non observed locations of the process tend to be closer to observations of the process, which implies that forecasting the values at the blind locations become possible and relies on the rate at which such gap is filled, namely m_N .

In the following, we will make the assumption that we can dispose of a consistent estimation procedure \hat{f}_N for f , such that there exists a decreasing sequence $r_N \rightarrow 0$, with is a rate of convergence of the spectral density when N goes to infinity, providing the controls on the following estimation errors

Assumption. *Preliminar consistent estimate for the spectral density*

- $\mathbb{E} \left[\left\| \hat{f}_N - f \right\|_{\infty}^2 \right]^{\frac{1}{2}} \leq r_N.$
- $\mathbb{E} \left[\left\| \hat{f}_N - f \right\|_{\infty}^4 \right]^{\frac{1}{4}} \leq r_N.$

Our aim is, observing only one realization X_{O_N} , to perform both estimation and prediction of any variable Z_B defined as a linear combination of the process taken at unobserved locations, i.e of the following form

$$Z_B = a_B^T X_B,$$

with $a_B \in \mathbb{R}^B$, $\|a_B\|_2 = 1$.

Hereafter, we will write, for sake of simplicity, extracted operators like block matrices even if they are of infinite size

$$\begin{aligned} \forall U, V \subset G, \mathcal{K}_{UV}(f) &= (\mathcal{K}(f)_{ij})_{i \in U, j \in V} \\ \forall V \subset G, \mathcal{K}_V(f) &= (\mathcal{K}(f)_{ij})_{i, j \in V} \end{aligned}$$

Recall (see for instance in [1]) that the best linear predictor of Z_B (which is also the best predictor in the Gaussian case) can be written as

$$\bar{Z}_B = P_{[X_{O_N}]}(f) Z_B := a_B^T \mathcal{K}_{BO_N}(f) (\mathcal{K}_{O_N}(f))^{-1} X_{O_N}.$$

$\mathcal{K}_{BO_N}(f)$ is the covariance between the observed process and the blind part, while $\mathcal{K}_{O_N}(f)$ corresponds to the covariance of the process restricted to the observed data points. Note that this projection term is well defined. Indeed since f is positive, $\mathcal{K}(f)$ is invertible, and therefore, $\mathcal{K}_{O_N}(f)$ is also invertible, as a principle minor. However, since f is unknown we can not use this as a direct estimator.

Then, remark that, asymptotically, in the sense $N \rightarrow +\infty$, we may observe the process at all locations, X_O . We thus can introduce the best linear prediction of Z_B knowing all the possible observations X_O as

$$\tilde{Z}_B := P_{[X_O]}(f) Z_B := a_B^T \mathcal{K}_{BO}(f) (\mathcal{K}_O(f))^{-1} X_O.$$

Finally, the blind forecasting problem can be formulated as a two step procedure mixing the estimation of the projector operator and the prediction using the estimated projector. It can be thus decomposed as follows

- Estimation step: estimate $P_{[X_{O_N}]}(f)$ by $\hat{P}_{[X_{O_N}]}(f) := P_{[X_{O_N}]}(\hat{f})$.
- Prediction step: build $\hat{Z}_B := P_{[X_{O_N}]}(\hat{f}) Z_B$.

Therefore, it seems natural, in order to analyze the forecast procedure, to consider an upper bound on the risk defined by

$$\mathbf{R}_N = \sup_{\substack{Z_B = a_B^T X_B \\ \|a_B\|_2 = 1}} \mathbb{E} \left[\left(Z_B - \hat{Z}_B \right)^2 \right]^{\frac{1}{2}}.$$

But we can see that this risk admits the following decomposition

$$\mathbf{R}_N = \sup_{\substack{Z_B = a_B^T X_B \\ \|a_B\|_2=1}} \left(\mathbb{E} \left[\left(Z_B - \tilde{Z}_B \right)^2 \right]^{\frac{1}{2}} + \mathbb{E} \left[\left(\tilde{Z}_B - \hat{Z}_B \right)^2 \right]^{\frac{1}{2}} \right).$$

Then, notice that the first term of this sum does not decrease to 0 when N goes to infinity. Actually it is an innovation type term. Therefore, we consider the upper bound

$$\begin{aligned} \mathbf{R}_N &\leq \sup_{\substack{Z_B = a_B^T X_B \\ \|a_B\|_2=1}} \mathbb{E} \left[\left(Z_B - \tilde{Z}_B \right)^2 \right]^{\frac{1}{2}} + \sup_{\substack{Z_B = a_B^T X_B \\ \|a_B\|_2=1}} \mathbb{E} \left[\left(\tilde{Z}_B - \bar{Z}_B \right)^2 \right]^{\frac{1}{2}} \\ &\quad + \sup_{\substack{Z_B = a_B^T X_B \\ \|a_B\|_2=1}} \mathbb{E} \left[\left(\bar{Z}_B - \hat{Z}_B \right)^2 \right]^{\frac{1}{2}} \\ &\leq \sup_{\substack{Z_B = a_B^T X_B \\ \|a_B\|_2=1}} \mathbb{E} \left[\left(Z_B - \tilde{Z}_B \right)^2 \right]^{\frac{1}{2}} + \mathcal{R}_N, \end{aligned}$$

where we have set

$$\mathcal{R}_N := \sup_{\substack{Z_B = a_B^T X_B \\ \|a_B\|_2=1}} \mathbb{E} \left[\left(\tilde{Z}_B - \bar{Z}_B \right)^2 \right]^{\frac{1}{2}} + \sup_{\substack{Z_B = a_B^T X_B \\ \|a_B\|_2=1}} \mathbb{E} \left[\left(\bar{Z}_B - \hat{Z}_B \right)^2 \right]^{\frac{1}{2}}.$$

The first term does not depend on the estimation procedure, hence our main issue is to compute the rate of convergence towards 0 of the two last terms of the previous sum. That is the reason why, \mathcal{R}_N plays the role of the estimation risk, that will be controlled in the whole paper.

2. PREDICTION OF A GRAPH PROCESS WITH INDEPENDENT OBSERVATIONS

In this section, we assume that another sample Y independent of X and drawn with the same distribution, is available. We will use Y to perform the estimation, and plug this estimation in order to predict Z_B .

More precisely, assume that for $p \geq 1$, \hat{f}_N is build using the sample Y observed on O_N and that X is observed on another subsample of the graph O_p . These nested collection of subgraphs fulfills the condition that $\frac{1}{m_p} = d_G(B, (G \setminus O_p))$ goes to infinity. In this first part, we thus have two independent asymptotics. The first one with respect to N controls how close the estimated spectral density will be from the

true one while the second with respect to p deals with the accuracy of the linear projector onto O_p .

We want to control the prediction of Z_B using the estimation \hat{f}_N .

To give an upper bound, we need to assume regularity for the spectral density f . First, we assume that the process has short range memory through the following assumption:

Assumption (A1). *There exists $M > 0$ such that*

$$\forall t \in \text{Sp}(A), m \leq f(t) \leq M.$$

Actually, we will assume that all the estimators should verify the same inequality, that is, for all $N \in \mathbb{N}$,

Assumption (A2).

$$\forall t \in \text{Sp}(A), m \leq \hat{f}_N(t) \leq M.$$

We will need another regularity assumption on f to control high-frequency behavior.

Assumption (A3). *The function $f(x) = \sum f_k x^k$ is analytic on a the compact disk $\bar{D}(0, 1)$ and verifies*

$$\sum k |f_k| \leq M.$$

Note that this assumption ensures that $\frac{1}{f}$ is absolutely convergent in 1 and that, if $\frac{1}{f}(x) = \sum (\frac{1}{f})_k x^k$,

$$\sum k |(\frac{1}{f})_k| \leq \frac{M}{m^2}.$$

The following theorem provides an upper bound for the risk $\mathcal{R}_{N,p}$

Theorem 1. *Assume that two independent samples are available. Under Assumptions (A1), (A2) and (A3), the risk admits the following upper bound*

$$\mathcal{R}_{N,p} \leq \frac{\sqrt{M}(m+M)}{m^2} r_N + \frac{M}{m^4} \left(\frac{M^{\frac{5}{2}}}{m} + M^{\frac{3}{2}} \right) m_p.$$

The prediction risk $\mathcal{R}_{N,p}$ is made of two separate terms which go to zero independently when p and N increase. The first term is entirely governed by the accuracy of the estimation of the spectral density. The second term depends on the decay of $\sum_{k \geq d_G(B, (G \setminus G_p))} \left| (\frac{1}{f})_k \right|$. This sum is the remaining term of a convergent sum and thus vanishes when p grows large. Note that using Assumption (A3), we obtain a bound in m_p . Adding more regularity on the function f of the type

$$\sum k^{2s} |(\frac{1}{f})_k| \leq \frac{M}{m^2}$$

for a given $s > 0$ would improve this rate by replacing m_p by m_p^s .

Proof. The proof consists in bounding the two terms in the decomposition of the risk $\mathcal{R}_{N,p}$.

The following lemmas give the rate of convergence of each of these terms.

Lemma 1. *The following upper bound holds:*

$$\sup_{\substack{z_B = a_B^T x_B \\ \|a_B\|_2 = 1}} \mathbb{E} \left[\left(\bar{Z}_B - \hat{Z}_B \right)^2 \right]^{\frac{1}{2}} \leq \frac{\sqrt{M}(m + M)}{m^2} r_N.$$

Lemma 2. *The following upper bound holds:*

$$\sup_{\substack{z_B = a_B^T x_B \\ \|a_B\|_2 = 1}} \mathbb{E} \left[\left(\bar{Z}_B - \tilde{Z}_B \right)^2 \right]^{\frac{1}{2}} \leq \frac{M}{m^4} \left(\frac{M^{\frac{5}{2}}}{m} + M^{\frac{3}{2}} \right) m_p.$$

The proofs of these Lemmas are postponed to the Appendix

□

3. THE BLIND CASE

Assume now that there is only one sample X available, observed on O_N . In this case, the two previous estimation steps are linked and the asymptotics between the estimation of the spectral density and the projection step must be carefully chosen.

In particular, we must pay a special attention to the behavior of the variance term to balance the two errors. Indeed, to perform the prediction, we will use the whole available sample for the estimation, but then chose a window $O_{p(N)} \subset O_N$, for a suitable $p(N)$ to make the prediction. Doing this, we may go beyond the dependency problem induced by the fact that the same sample has to be used for both estimation and filtering.

The following theorem provides the rate of convergence of the error term.

Theorem 2. *Assume that $\hat{f}_{p(N)}$ is built with the observation sample X and that Assumptions [A1] and [A2] holds. The risk admits the following upper bound*

$$\mathcal{R}_{N,p(N)} \leq \frac{m + M}{m^2} \mathbb{E} \left[\left(\sum_{i \in O_{p(N)}} X_i^2 \right)^2 \right]^{\frac{1}{4}} r_N + \frac{M}{m^4} \left(\frac{M^{\frac{5}{2}}}{m} + M^{\frac{3}{2}} \right) m_{p(N)}.$$

We point out that we obtain two terms alike those in Theorem 1. The main difference comes from the extra term

$$\mathbb{E} \left[\left(\sum_{i \in O_{p(N)}} X_i^2 \right)^2 \right]^{\frac{1}{4}},$$

which corresponds to the price to pay to use the same observation sample. Using $\mathbb{E}[X_i^2 X_j^2] \leq 3\mathbb{E}[X_i^2]\mathbb{E}[X_j^2]$, leads to the following upper bound

$$\begin{aligned} \mathbb{E} \left[\left(\sum_{i \in O_{p(N)}} X_i^2 \right)^2 \right]^{\frac{1}{4}} &\leq 3 \sum_{i \in O_{p(N)}} (\mathbb{E}[X_i^2])^2 \\ &\leq (\#O_{p(N)})^2, \end{aligned}$$

where $\#$ denotes the cardinal of a set. Finally, we see that the two error terms are linked in an opposite way such that the window parameter $p(N)$ must be chosen to balance the bias and the variance by minimizing with respect to $p(N)$ the quantity

$$(1) \quad \frac{m+M}{m^2} \sqrt{\#O_{p(N)} r_N} + \frac{M}{m^4} \left(\frac{M^{\frac{5}{2}}}{m} + M^{\frac{3}{2}} \right) m_{p(N)}.$$

For instance, consider the special case of a process defined on \mathbb{Z}^2 . Assume that B is a finite subset of \mathbb{Z}^2 , and that we have at hand the following sequence of nested sub graphs $G_N = [-N, N]^2$. Then we get the following approximations for the quantities of Equation 1, for some constants C, k ,

- $\#O_N$ is of order $4N^2$
- r_N is of order $\frac{C}{N}$ (see for instance [4])
- m_p is of order $\frac{1}{p-k}$.

Then minimizing (1) implies minimizing

$$C_1 \frac{p(N)}{N} + C_2 \frac{1}{p(N) - k},$$

which is achieved for $p(N) \approx \sqrt{N}$. Therefore the error is such that

$$\mathcal{R}_{N,p(N)} = O\left(\frac{1}{\sqrt{N}}\right).$$

Here, the blind case leads with this method to an important loss since the error in the case of the independent sample would have been of order $\frac{1}{N}$. A lower bound would be necessary to fully clarify this result, however obtaining such a bound seems a very difficult task which falls beyond the scope of this paper. Nevertheless, to our

knowledge, the blind prediction of a graph indexed random process has never been tackled before.

Proof. The only thing which remains to be calculated is given by the following lemma.

Lemma 3. *The following upper bound holds:*

$$\sup_{\substack{Z_B = a_B^T X_B \\ \|a_B\|_2 = 1}} \mathbb{E} \left[\left(\bar{Z}_B - \hat{Z}_B \right)^2 \right]^{\frac{1}{2}} \leq \frac{m+M}{m^2} \mathbb{E} \left[\left(\sum_{i \in O_p(N)} X_i^2 \right)^2 \right]^{\frac{1}{4}} r_N.$$

The proof is postponed in Appendix

□

4. APPENDIX

Proof. of Lemma 1

First, denote

$$A_{N,p} \left(f, \hat{f}_N \right) = \mathcal{K}_{BO_p}(\hat{f}_N) \left(\mathcal{K}_{O_p}(\hat{f}_N) \right)^{-1} - \mathcal{K}_{BO_p}(f) \left(\mathcal{K}_{O_p}(f) \right)^{-1},$$

and note that, $\mathbb{P}_X \otimes \mathbb{P}_Y$ -a.s.,

$$\left(\hat{Z}_B - \bar{Z}_B \right)^2 = a_B^T \left(A_{N,p} \left(f, \hat{f}_N \right) \right)^T X_{O_p} X_{O_p}^T A_{N,p} \left(f, \hat{f}_N \right) a_B$$

Then, notice that

$$\begin{aligned} & \sup_{\substack{Z_B = a_B^T X_B \\ \|a_B\|_2 = 1}} \mathbb{E} \left[\left(\bar{Z}_B - \hat{Z}_B \right)^2 \right]^{\frac{1}{2}} \\ & \leq \left(\sup_{\substack{\text{supp}(a_B) \subset B \\ \|a_B\|_2 = 1}} \mathbb{E}_{\mathbb{P}_Y} \left[a_B^T \left(A_{N,p} \left(f, \hat{f}_p \right) \right)^T \mathcal{K}_{O_p}(f) A_{N,p} \left(f, \hat{f}_N \right) a_B \right] \right)^{\frac{1}{2}} \\ & \leq \mathbb{E}_{\mathbb{P}_Y} \left[\sup_{\substack{\text{supp}(a_B) \subset B \\ \|a_B\|_2 = 1}} a_B^T \left(A_{N,p} \left(f, \hat{f}_N \right) \right)^T \mathcal{K}_{O_p}(f) A_{N,p} \left(f, \hat{f}_N \right) a_B \right]^{\frac{1}{2}}. \end{aligned}$$

Thus,

$$\sup_{\substack{Z_B = a_B^T X_B \\ \|a_B\|_2 = 1}} \mathbb{E} \left[\left(\bar{Z}_B - \hat{Z}_B \right)^2 \right]^{\frac{1}{2}} \leq \mathbb{E}_{\mathbb{P}_Y} \left[\left\| \left(A_{N,p} \left(f, \hat{f}_N \right) \right)^T \mathcal{K}_{O_p}(f) A_{N,p} \left(f, \hat{f}_N \right) \right\|_{2,op} \right]^{\frac{1}{2}}.$$

Furthermore, it holds, \mathbb{P}_Y -a.s., that

$$\left\| \left(A_{N,p} \left(f, \hat{f}_N \right) \right)^T \mathcal{K}_{O_p}(f) A_{N,p} \left(f, \hat{f}_N \right) \right\|_{2,op} \leq \left\| A_{N,p} \left(f, \hat{f}_N \right) \right\|_{2,op}^2 \left\| \mathcal{K}_{O_p}(f) \right\|_{2,op}$$

But,

$$\begin{aligned} \left\| A_{N,p} \left(f, \hat{f}_N \right) \right\|_{2,op} &\leq \left\| \mathcal{K}_{BO_N}(\hat{f}_N) \left(\mathcal{K}_{O_p}(\hat{f}_N) \right)^{-1} - \mathcal{K}_{BO_p}(f) \left(\mathcal{K}_{O_p}(\hat{f}_N) \right)^{-1} \right\|_{2,op} \\ &\quad + \left\| \mathcal{K}_{BO_p}(f) \left(\mathcal{K}_{O_p}(\hat{f}_N) \right)^{-1} - \mathcal{K}_{BO_p}(f) \left(\mathcal{K}_{O_p}(f) \right)^{-1} \right\|_{2,op} \\ &\leq \left\| \left(\mathcal{K}_{O_p}(\hat{f}_N) \right)^{-1} \right\|_{2,op} \left\| \mathcal{K}_{BO_p}(\hat{f}_N) - \mathcal{K}_{BO_p}(f) \right\|_{2,op} + \left\| \mathcal{K}_{BO_p}(f) \right\|_{2,op} \\ &\quad \times \left\| \left(\mathcal{K}_{O_p}(f) \right)^{-1} \right\|_{2,op} \left\| \left(\mathcal{K}_{O_p}(\hat{f}_N) \right)^{-1} \right\|_{2,op} \left\| \mathcal{K}_{O_p}(\hat{f}_N) - \mathcal{K}_{O_p}(f) \right\|_{2,op} \\ &\leq \frac{1}{m} \left\| f - \hat{f}_N \right\|_{\infty} + \frac{M}{m^2} \left\| f - \hat{f}_N \right\|_{\infty} \end{aligned}$$

Here we used the inequality

$$\left\| \mathcal{K}(f) \right\|_{2,op} \leq \left\| f \right\|_{\infty}$$

We get

$$\begin{aligned} \sup_{\substack{Z_B = a_B^T X_B \\ \|a_B\|_2 = 1}} \mathbb{E} \left[\left(\bar{Z}_B - \hat{Z}_B \right)^2 \right]^{\frac{1}{2}} &\leq \frac{\sqrt{M}(m+M)}{m^2} \mathbb{E}_{\mathbb{P}_Y} \left[\left\| f - \hat{f}_N \right\|_{\infty}^2 \right]^{\frac{1}{2}} \\ &\leq \frac{\sqrt{M}(m+M)}{m^2} r_N \end{aligned}$$

□

Proof. of Lemma 2

First, define for all $A \subset G$, the operator p_A by

$$\forall i, j \in G, (p_A)_{ij} = \mathbb{1}_{i \in A} \mathbb{1}_{i=j}.$$

To compute the rate of convergence of the bias term $\bar{Z}_B - \tilde{Z}_B$, we can compute directly,

$$\begin{aligned}
& \sup_{\substack{Z_B = a_B^T X_B \\ \|a_B\|_2 = 1}} \mathbb{E} \left[\left(\bar{Z}_B - \tilde{Z}_B \right)^2 \right]^{\frac{1}{2}} \\
& \leq \left\| \mathcal{K}_{BO_p}(f) (\mathcal{K}_{O_p}(f))^{-1} p_{O_p} - K_{BO}(f) (\mathcal{K}_O(f))^{-1} \right\|_{2,op} \|\mathcal{K}_O(f)\|_{2,op}^{\frac{1}{2}}
\end{aligned}$$

Note that, since $B \cup O = G$, we have immediately that $(\mathcal{K}_{B \cup O}(f))^{-1} = \mathcal{K}(\frac{1}{f})$. Using a Schur decomposition, we get

$$\begin{aligned}
& \sup_{\substack{Z_B = a_B^T X_B \\ \|a_B\|_2 = 1}} \mathbb{E} \left[\left(\bar{Z}_B - \tilde{Z}_B \right)^2 \right]^{\frac{1}{2}} \\
& \leq \sqrt{M} \left\| \mathcal{K}_{BO_p}(f) (\mathcal{K}_{O_p}(f))^{-1} + \left(\mathcal{K}_B(\frac{1}{f}) \right)^{-1} \mathcal{K}_{BO}(\frac{1}{f}) \right\|_{2,op} \\
& \leq \sqrt{M} \left\| \mathcal{K}_{BO_p}(f) (\mathcal{K}_{O_p}(f))^{-1} + \left(\mathcal{K}_B(\frac{1}{f}) \right)^{-1} \mathcal{K}_{BO_p}(\frac{1}{f}) \right\|_{2,op} \\
& \quad + \sqrt{M} \left\| \left(\mathcal{K}_B(\frac{1}{f}) \right)^{-1} \right\|_{2,op} \left\| \mathcal{K}_{B(O \setminus O_p)}(\frac{1}{f}) \right\|_{2,op} \\
& \leq \sqrt{M} \left\| \mathcal{K}_B(\frac{1}{f}) \mathcal{K}_{BO_p}(f) + \mathcal{K}_{BO_N}(\frac{1}{f}) \mathcal{K}_{O_p}(f) \right\|_{2,op} \\
& \quad \times \left\| (\mathcal{K}_{O_p}(f))^{-1} \right\|_{2,op} \left\| \left(\mathcal{K}_B(\frac{1}{f}) \right)^{-1} \right\|_{2,op} + M^{\frac{3}{2}} \left\| \mathcal{K}_{B(O \setminus O_p)}(\frac{1}{f}) \right\|_{2,op} \\
& \leq \frac{M^{\frac{3}{2}}}{m} \left\| -\mathcal{K}_{B(O \setminus O_p)}(\frac{1}{f}) \mathcal{K}_{(O \setminus O_p)O_p}(f) \right\|_{2,op} \\
& \quad + M^{\frac{3}{2}} \left\| \mathcal{K}_{B(O \setminus O_p)}(\frac{1}{f}) \right\|_{2,op}
\end{aligned}$$

This leads to

$$(2) \quad \sup_{\substack{Z_B = a_B^T X_B \\ \|a_B\|_2 = 1}} \mathbb{E} \left[(\bar{Z}_B - \tilde{Z}_B)^2 \right]^{\frac{1}{2}} \leq \left(\frac{M^{\frac{5}{2}}}{m} + M^{\frac{3}{2}} \right) \left\| \mathcal{K}_{B(O \setminus O_p)} \left(\frac{1}{f} \right) \right\|_{2,op}$$

Moreover, we can write

$$\begin{aligned} \left\| \mathcal{K}_{B(O \setminus O_p)} \left(\frac{1}{f} \right) \right\|_{2,op} &\leq \sum_{k \geq 0} \left| \left(\frac{1}{f} \right)_k \right| \left\| (A^k)_{B(O \setminus O_p)} \right\|_{2,op} \\ &\leq \sum_{k \geq d_G(B, (O \setminus O_p))} \left| \left(\frac{1}{f} \right)_k \right| \\ &\leq \sum_{k \geq d_G(B, (G \setminus G_p))} \left| \left(\frac{1}{f} \right)_k \right| \\ &\leq \frac{1}{d_G(B, (G \setminus G_p))} \sum_{k \geq d_G(B, (G \setminus G_p))} k \left| \left(\frac{1}{f} \right)_k \right| \\ &\leq \frac{M}{m^2} m_p \end{aligned}$$

□

Proof. of Lemma 3

For sake of simplicity, let us denote O_p instead of $O_{p(N)}$ in the whole proof.
We still denote

$$A_{N,p} \left(f, \hat{f}_N \right) = \mathcal{K}_{BO_p}(\hat{f}_N) \left(\mathcal{K}_{O_p}(\hat{f}_N) \right)^{-1} - \mathcal{K}_{BO_p}(f) \left(\mathcal{K}_{O_p}(f) \right)^{-1},$$

Then, notice that

$$\begin{aligned}
& \sup_{\substack{Z_B = a_B^T X_B \\ \|a_B\|_2 = 1}} \mathbb{E} \left[\left(\bar{Z}_B - \hat{Z}_B \right)^2 \right]^{\frac{1}{2}} \\
& \leq \left(\sup_{\substack{\text{supp}(a_B) \subset B \\ \|a_B\|_2 = 1}} \mathbb{E} \left[\text{Tr} \left(a_B^T \left(A_{N,p} \left(f, \hat{f}_N \right) \right)^T X_{O_p} X_{O_p}^T A_N \left(f, \hat{f}_N \right) a_B \right) \right] \right)^{\frac{1}{2}} \\
& \leq \left(\sup_{\substack{\text{supp}(a_B) \subset B \\ \|a_B\|_2 = 1}} \mathbb{E} \left[\text{Tr} \left(X_{O_p} X_{O_p}^T A_{N,p} \left(f, \hat{f}_N \right) a_B a_B^T \left(A_{N,p} \left(f, \hat{f}_N \right) \right)^T \right) \right] \right)^{\frac{1}{2}}
\end{aligned}$$

Applying Cauchy Schwartz inequality, we get

$$\begin{aligned}
& \sup_{\substack{Z_B = a_B^T X_B \\ \|a_B\|_2 = 1}} \mathbb{E} \left[\left(\bar{Z}_B - \hat{Z}_B \right)^2 \right]^{\frac{1}{2}} \\
& \leq \left(\mathbb{E} \left[\text{Tr} \left(\left(X_{O_p} X_{O_p}^T \right)^2 \right) \right] \right. \\
& \quad \times \left. \sup_{\substack{\text{supp}(a_B) \subset B \\ \|a_B\|_2 = 1}} \mathbb{E} \left[\text{Tr} \left(\left(A_{N,p} \left(f, \hat{f}_N \right) a_B a_B^T \left(A_{N,p} \left(f, \hat{f}_N \right) \right)^T \right)^2 \right) \right] \right)^{\frac{1}{4}}
\end{aligned}$$

But, on the one hand, we have

$$\begin{aligned}
\mathbb{E} \left[\text{Tr} \left(\left(X_{O_p} X_{O_p}^T \right)^2 \right) \right] &= \mathbb{E} \left[\text{Tr} \left(\left(X_{O_p}^T X_{O_p} \right)^2 \right) \right] \\
&= \mathbb{E} \left[\left(\sum_{i \in O_p} X_i^2 \right)^2 \right]
\end{aligned}$$

And on the other hand, we can write

$$\begin{aligned}
& \sup_{\substack{\text{supp}(a_B) \subset B \\ \|a_B\|_2=1}} \mathbb{E} \left[\text{Tr} \left(\left(A_N(f, \hat{f}_N) a_B a_B^T (A_{N,p}(f, \hat{f}_N))^T \right)^2 \right) \right] \\
& \leq \sup_{\substack{\text{supp}(a_B, b_B) \subset B \\ \|a_B\|_2=\|b_B\|_2=1}} \mathbb{E} \left[\text{Tr} \left(a_B^T (A_{N,p}(f, \hat{f}_N))^T A_{N,p}(f, \hat{f}_N) \right. \right. \\
& \quad \left. \left. \times b_B b_B^T (A_{N,p}(f, \hat{f}_N))^T A_{N,p}(f, \hat{f}_N) a_B \right) \right] \\
& \leq \sup_{\substack{\text{supp}(b_B) \subset B \\ \|b_B\|_2=1}} \mathbb{E} \left[\|A_{N,p}(f, \hat{f}_N)\|_{2,op}^4 \|b_B b_B^T\|_{2,op} \right] \\
& \leq \mathbb{E} \left[\|A_{N,p}(f, \hat{f}_N)\|_{2,op}^4 \right] \\
& \leq \left(\frac{m+M}{m^2} \right)^4 \mathbb{E} \left[\|f - \hat{f}_N\|_\infty^4 \right]
\end{aligned}$$

Thus,

$$\sup_{\substack{z_B = a_B^T x_B \\ \|a_B\|_2=1}} \mathbb{E} \left[\left(\bar{Z}_B - \hat{Z}_B \right)^2 \right]^{\frac{1}{2}} \leq \frac{m+M}{m^2} \mathbb{E} \left[\left(\sum_{i \in O_p} X_i^2 \right)^2 \right]^{\frac{1}{4}} r_N$$

□

REFERENCES

- [1]
- [2] R. Azencott and D. Dacunha-Castelle. *Series of irregular observations*. Applied Probability. A Series of the Applied Probability Trust. Springer-Verlag, New York, 1986. Forecasting and model building.
- [3] N. A. C. Cressie. *Statistics for spatial data*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York, 1993. Revised reprint of the 1991 edition, A Wiley-Interscience Publication.
- [4] R. Dahlhaus and H. Künsch. Edge effects and efficient parameter estimation for stationary random fields. *Biometrika*, 74(4):877–882, 1987.
- [5] J. N. Darroch, S. L. Lauritzen, and T. P. Speed. Markov fields and log-linear interaction models for contingency tables. *Ann. Statist.*, 8(3):522–539, 1980.

- [6] T. Espinasse, F. Gamboa, and J.-M. Loubes. Estimation Error for Blind Gaussian Time Series Prediction. *Maths. Methods in Statistics*, 20(3):206–223, 2011.
- [7] T. Espinasse, F. Gamboa, and J.-M. Loubes. Parametric estimation for gaussian fields indexed by graphs. *Probability Theory and Related Fields*, pages 1–39, 2012.
- [8] D. G. Krige. A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, 52(6):119–139, Dec. 1951.
- [9] L. C. M., L. S., R. X., H. F., and J. B. *Colloque Arcéométrie GMPCA et Centre Européen d’Archéométrie*.
- [10] N. Verzelen. Adaptive estimation of stationary Gaussian fields. *Ann. Statist.*, 38(3):1363–1402, 2010.
- [11] N. Verzelen and F. Villers. Tests for Gaussian graphical models. *Comput. Statist. Data Anal.*, 53(5):1894–1905, 2009.